

# The Toxicity Data Landscape for Environmental Chemicals

Richard Judson(1), Ann Richard(1), David J. Dix(1), Keith Houck(1), Matthew Martin(1), Robert Kavlock(1), Vicki Dellarco(2), Todd Holderman(3), Shirlee Tan(4), Thomas Carpenter(5), Edwin Smith(6)

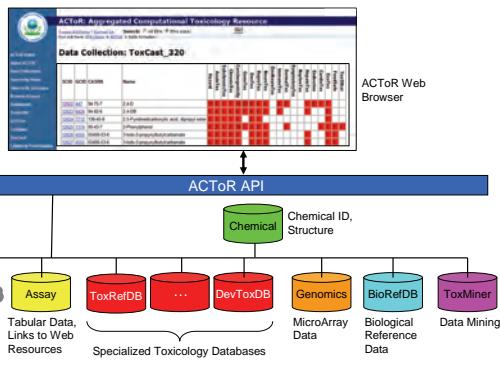
(1)NCCT/ORD, USEPA, RTP NC, USA, (2)OPP/OPPTS, USEPA, Arlington VA, USA, (3)OPPT/OPPTS, USEPA, Washington DC, USA, (4)OSCP/OPPTS, USEPA, Washington DC, USA, (5)OW/OGWDW, USEPA, Washington DC, USA, (6)GLNPO, USEPA, Chicago IL, USA

**Abstract:** Chemical toxicity information is widely dispersed across multiple sources, which hampers the compilation of data for large sets of chemicals. To address this issue, we are developing ACToR (Aggregated Computational Toxicology Resource), which combines information for ~500,000 chemicals from >200 sources including the EPA, NIH, FDA, and corresponding agencies in Canada, Europe and Japan. The database holds chemical structure; physico-chemical properties; biochemical and cell-based assay data; tabular in vivo data; summary toxicology calls; and URLs to toxicology summaries. An important application is to provide a source of consistent in vivo toxicology data for use in toxicity screening and prioritization efforts.

Tens of thousands of chemicals are in common use but only a fraction have had significant toxicological evaluation due to the expense of animal testing. An alternative screening and prioritization approach uses *in vitro* assay "signatures" that can predict *in vivo* phenotypes. This approach can rapidly screen many chemicals and prioritize "positive" chemicals for further evaluation. ACToR enables the consistent compilation of *in vivo* toxicology data for many chemicals, which is required for screening signature development and validation.

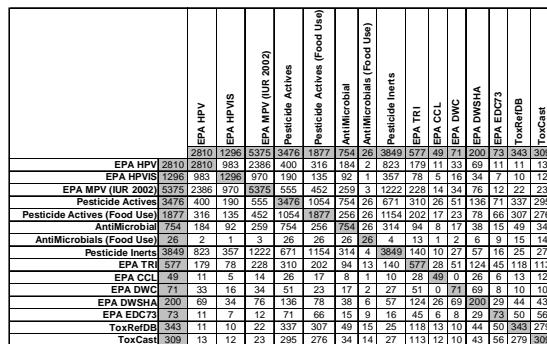
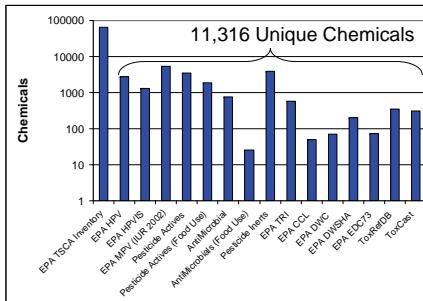
We show results for a set of ~11,000 environmental chemicals to be analyzed as part of the EPA ToxCast™ screening and prioritization program. These include High and Medium Production Volume chemicals (HPVs and MPVs), pesticide active and inert ingredients and drinking water contaminants.

ACToR is based on open source technology and will allow full access both EPA-wide and to the public.

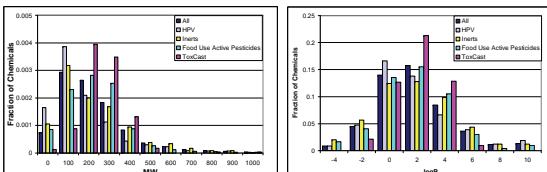


## Candidate Chemicals for Screening and Prioritization

ACToR is aggregating data for environmental chemicals to characterize data coverage for the ToxCast screening and prioritization project



## Sample Physico-Chemical Distributions



Many types of data are included in order to gain the widest possible coverage.

Data Type	Notes
Physicochemical Properties	Calculated from Chemical Structure
Biochemical (In Vitro or Cell-based) Assay Data	In vitro High Throughput Screening (HTS) data
In Vivo Toxicology Assay Data (Primary Tabular)	Tabulated data from primary source, e.g. ToxRefDB
In Vivo Toxicology Assay Data (Secondary Tabular)	Tabulated data from secondary source, e.g. IRIS
In Vivo Toxicology Text Reports via URL	Pointers to text reports on the Web, e.g. EPA CARC Human Cancer Categories
Regulatory Listings	From EPA and U.S. State Websites
Chemical Categories	From EPA and OECD
Phenotypes	Assigned from Study Types

## Data Types in ACToR

Many types of data are included in order to gain the widest possible coverage.

Data Type	Notes
ATSDR (multiple)	Health Canada
CalEPA	EPA SRS
CERCLA	EPA ITRI
CERHR	EPA CCL
CPDB	EPA EFED
Danish EPA	EPA IRIS
DSSTox	KDB
EINECS	RAIS (multiple)
Environment Canada	MHLW Japan
EU ToxCast	ITER TERA
FDA	ToxCast
HMIS	MTI Japan
NCGC	TOXNET (multiple)
OECD	WHO
NIOSH	INCHM
ToxRefDB	INNAD
WHO	OECD (multiple)

## Major Data Sources

Data Sources	Health Canada	EPA SRS	INCHM (multiple)	NTP (multiple)
CalEPA	EPA CARC	EPA ITRI	INNAD	OECD (multiple)
CERCLA	EPA CCL	EPA IRIS	PUBCHEM	
CERHR	EPA EFED	EPA IUR	KDB	RAIS (multiple)
CPDB	EPA HPVIS	ESIS (multiple)	MHLW Japan	ToxCast
Danish EPA	EXTOXNET	FDA CDER	MTI Japan	TOXNET (multiple)
DSSTox	EPA PFATE	FDA (multiple)	NCGC	ToxRefDB
EINECS	EPA RBC	HERA	NCI	WHO
Environment Canada	EPA RSEI	HMDB	NIOSH	

## ACToR Database Statistics

Data Sources	235
Unique Chemicals	525,000
Chemicals with structure	414,000
Assays	1702
Assay Data Points	7 M

## Data Coverage Summary for the 11,316 Chemicals

Data is categorized by type and whether the source breaks out data by phenotype or whether data from multiple phenotypes are collected in a single report which will have to be manually extracted.

## Chemical Coverage by Phenotype and Data Type

Name	Total	% Coverage
Carcinogenicity - Primary-Tabular	690	6.1%
Carcinogenicity - URL	1799	15.5%
Carcinogenicity - Secondary	728	6.4%
Carcinogenicity - All	2561	22.6%
GeneTox - Primary-Tabular	1238	10.9%
GeneTox - URL	221	2.0%
GeneTox - All	1363	12.0%
DevTox - Primary-Tabular	378	3.3%
DevTox - URL	2365	20.5%
DevTox - Secondary-Tabular	221	2.0%
DevTox - All	2519	22.3%
Reprotox - Primary-Tabular	294	2.6%
Reprotox - URL	69	0.6%
Reprotox - All	404	3.6%
Multiple - URL	4349	38.4%
Multiple - All	4435	39.2%

This poster does not necessarily reflect EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.



COMPUTATIONAL  
TOXICOLOGY